

sgRNAs9 软件图形用户界面开发及应用

赵长志¹, 张懿³, 李广磊¹, 陈济良², 李京津¹, 任瑞敏¹, 倪攀¹,
赵书红¹, 谢胜松^{1,2}

1. 华中农业大学, 农业动物遗传育种与繁殖教育部重点实验室, 武汉 430070;

2. 华中农业大学, 动物医学院动物生理生化教研室, 武汉 430070;

3. 中国人民解放军第 161 医院妇产科, 武汉 430010

摘要: CRISPR/Cas9 系统介导的基因组编辑技术是新一代功能强大的基因修饰技术。然而, 脱靶效应(Off-target effects)是目前 CRISPR/Cas9 技术面临的重大问题。因此, 设计脱靶风险低的 sgRNA(single guide RNA)就成为关键。sgRNAs9 是一款专门用于 sgRNA 设计和评估脱靶效应的软件包。针对其核心运行程序, 我们利用 Java 程序语言开发了其图形用户界面。此外, 依据脱靶位点碱基数和 sgRNA 种子序列(seed sequences)的特异性, 通过设置不同的风险等级对 sgRNA 的脱靶效应进行评估。随后, 利用此软件设计了 34 124 条靶向人、小鼠、大鼠、猪和鸡中共 4691 个 microRNA(miRNA)前体的 sgRNAs。此外, 随机挑选了一个靶向人 miR-206 前体的 sgRNA 进行脱靶效应评估和验证。结果发现, sgRNAs9 软件人机交互界面友好, 大多数 miRNA 前体可通过该软件寻找到 sgRNA, 且他们的 GC%含量范围集中于 40%~60%。利用 sgRNAs9 软件设计的靶向 miR-206 的 sgRNA, 其基因组编辑活性和脱靶位点可被实验验证。本研究表明 sgRNAs9 图形用户界面软件能针对任意物种设计特异性的 sgRNA, 其可在 BioTools(<http://www.biotoools.com/>)网站下载。

关键词: CRISPR/Cas9; sgRNAs9 软件; 基因组编辑; miRNA; 计算机辅助设计

Development of a graphical user interface for sgRNAs9 and its application

Changzhi Zhao¹, Yi Zhang³, Guanglei Li¹, Jiliang Chen², JingJin Li¹, Ruimin Ren¹,
Pan Ni¹, Shuhong Zhao¹, Shengsong Xie^{1,2}

1. Key Lab of Agricultural Animal Genetics, Breeding, and Reproduction of Ministry of Education, Huazhong Agricultural University, Wuhan 430070, China

2. Department of Animal Physiology and Biochemistry, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan 430070, China;

3. Department of Obstetrics and Gynecology, No.161 Hospital of PLA, Wuhan 430010, China

Abstract: The CRISPR/Cas9 genome editing technique is a powerful tool for researchers. However, off-target effects of the Cas9 nuclease activity is a recurrent concern of the CRISPR system. Thus, designing sgRNA (single

收稿日期: 2015-06-21; 修回日期: 2015-08-29

基金项目: 国家转基因重大专项项目(编号: 2014ZX08006-005)、广东省分子与细胞工程育种团队(编号: 2011A020102003)、国家自然科学基金项目(编号: 31301226)和中央高校基本科研业务费专项资金资助项目(批准号: 2662015BQ004 和 2662015BQ005)资助

作者简介: 赵长志, 博士研究生, 研究方向: 动物遗传育种与繁殖。Email: 935618112@qq.com.

通讯作者: 谢胜松, 博士, 副研究员, 研究方向: 动物功能基因组与疾病诊治。Email: sxxie@mail.hzau.edu.cn.

DOI: 10.16288/j.ycz.15-292

网络出版时间: 2015-9-8 8:41:28

URL: <http://www.cnki.net/kcms/detail/11.1913.R.20150908.0841.002.html>

guide RNA) with minimal off-target effects is very important. sgRNAs9 is a software package, which can be used to design sgRNA and to evaluate potential off-target cleavage sites. In this study, a graphical user interface for sgRNAs9 was developed using the Java programming language. In addition, off-target effect for sgRNAs was evaluated according to mismatched number and “seed sequence” specification. Moreover, sgRNAs9 software was used to design 34 124 sgRNAs, which can target 4691 microRNA (miRNA) precursors from human, mouse, rat, pig, and chicken. In particular, the off-target effect of a sgRNA targeting to human miR-206 precursor was analyzed, and the on/off-target activity of this sgRNA was validated by T7E1 assay *in vitro*. Taken together, these data showed that the interface can simplify the usage of the sgRNAs9 program, which can be used to design sgRNAs for the majority of miRNA precursors. We also found that the GC% of those sgRNAs ranged from 40% to 60%. In summary, the sgRNAs9 software can be easily used to design sgRNA with minimal off-target effects for any species. The software can be downloaded from BiooTools website (<http://www.biootools.com/>).

Keywords: CRISPR/Cas9; sgRNAs9 software; genome editing; miRNA; computer-aided design

基于 CRISPR(Clustered regularly interspaced short palindromic repeats)/Cas9 系统的基因组编辑技术来源于细菌获得性免疫, 是一种由 RNA 引导 Cas9 核酸酶对靶基因进行修饰的新技术^[1,2]。CRISPR/Cas9 技术包含两部分: 一是 sgRNA(Single guide RNA), 通过碱基互补配对与 DNA 特异结合; 其次是 Cas9 核酸酶, 可与具有下游前间区序列邻近基序(Protospacer adjacent motif, PAM)的特定双链 DNA 结合并进行切割。该技术可在特定位点实现非同源末端连接(Non-homologous end joining, NHEJ)或同源重组(Homologous recombination, HR), 目前已成为继锌指核酸酶(Zinc-finger nucleases, ZFNs)和类转录激活因子效应物核酸酶(Transcription activator-like effector nuclease, TALEN)技术之后的第三代基因组编辑技术^[3-9]。该技术已成功在包括食蟹猴(*Cynomolgus monkey*)^[10]、小鼠(*Mus musculus*)^[11]、大鼠(*Rattus norvegicus*)^[12]、猪(*Sus scrofa*)^[13]、斑马鱼(*Zebrafish*)^[14]、水稻(*Oryza sativa*)、高粱(*Sorghum bicolor*)^[15]和小麦(*Triticum aestivum*)^[16]等多个物种中实现基因敲除, 并开始用于破坏人类免疫缺陷病毒(Human immunodeficiency virus, HIV)和乙型肝炎病毒(Hepatitis B virus, HBV)^[17-19]。除了用于基因敲除外, 它还可激活基因表达或替代 RNA 干扰(RNA interference, RNAi)技术抑制基因表达^[20,21]。相信在未来数年内, 这项新的基因组编辑技术必将成为每个分子生物学实验室必备的基因功能研究利器。然而多项研究报告指出, CRISPR/Cas9 系统介导的基因组编辑技术存在严重的脱靶副作用^[22,23]。Pattanayak 等^[24]利用高通量测序方法, 对 sgRNA 的特异性进行了检测, 发现其特异

性与临近 PAM 的 8~12 个碱基对组成有关, 即种子序列(Seed sequence)。因此, 预测 sgRNA 的脱靶位点时, 除了考察其错配碱基数外, 还需确定错配碱基是否位于种子序列区。尽管 CRISPR/Cas9 技术实验操作方便, 基因组编辑效率相对较高, 但脱靶问题一直制约其发展。除了选择合适类型的 Cas9 核酸酶外, 选择高效且特异的 sgRNA 也是实验成功的关键性因素之一。

目前, 已有 20 多款针对 CRISPR/Cas9 系统的 sgRNA 设计和/或脱靶效应评估软件被开发。这些软件可分为在线设计和单机版, 在线设计软件人机交互界面友好, 如 Optimized CRISPR Design^[25]、E-CRISP^[26]、CRISPR-P^[27]、CHOPCHOP^[28]和 CRISPR-direct^[29]等。而对于脱靶效应评估, 仅限于对提供有该物种基因组的 sgRNA 预测脱靶位点, 例如 Optimized CRISPR Design 仅针对人、小鼠和大鼠等 15 个物种设计 sgRNA, 而 CRISPR-P 软件含有 33 种植物的基因组, 主要用于设计植物的 sgRNA。单机版有 CasOT^[30]、SSFinder^[31]和 sgRNAs9^[32]等。相对而言, 单机版软件均为命令行工具(command-line tool), 需要用户具备一定的生物信息学基础, 安装与使用难度大。但最大优点是, 单机版软件可针对任意物种设计 sgRNA, 并可用于设计大规模全基因组 sgRNA 文库。

相对于其他单机版软件, 本实验室前期开发的 sgRNAs9 软件功能丰富^[32], 其包含有 8 个功能不同的小程序, 最新版本号为 V3.0.5(www.bioo-tools.com)。该软件包核心运行程序为“sgRNAs9.pl”, 用户可设置的参数有: (1) sgRNA 的长度, 限定范围

为 17~20 nt; (2) sgRNA 的 GC%含量,可限定最大值和最小值; (3)可限定靶标在 DNA 链的位置,如 DNA 的正义链、反义链或双链; (4) 设计 2 种类型的 sgRNA: 单个 sgRNA 或“paired-gRNA”。如选择设计“paired-gRNA”,用户可限定 2 个 sgRNA 之间距离最大值和最小值。对于 sgRNA 设计,其考察的 PAM 类型为“NGG”(N 代表 ATCG),而评估 sgRNA 的脱靶位点时,综合考察了两种 PAM 类型,即“NGG”和“NAG”,预测脱靶除了最大允许含 5 个碱基错配外,还特别考察了错配碱基是否位于种子序列。该软件由 Perl 语言编写,需在命令行终端通过输入命令调用,因此操作起来较困难。为了方便用户使用,本文针对 sgRNAs9 软件的核心程序,利用 Java 编程语言开发了其图形用户界面,并对 sgRNA 的脱靶风险评估算法和结果输出文件进行了优化。利用 sgRNAs9 软件,本文还对 miRBase 数据库^[33] (<http://www.mirbase.org/>)收录的人、小鼠、大鼠、猪和鸡的 miRNA 前体设计了 sgRNA,并对这些 sgRNA 的特征进行分析,由此构建一个针对 5 个物种所有 miRNA 前体的 sgRNA 数据集,最后对靶向 hsa-miR-206 成熟序列的一个 sgRNA 进行脱靶效应评估,并通过 T7E1 酶切实验检测了该 sgRNA 的打靶活性和预测的脱靶位点。目前,sgRNAs9 图形用户界面软件,全称为“CRISPR/Cas9 系统介导的基因组编辑技术 sgRNA 设计和脱靶效应评估软件”,已获得了中国版权保护中心颁发的计算机软件著作权登记证书(登记号: 2015SR170234)。该软件图形用户界面的开发,必将方便广大科研工作者使用。

1 材料和方法

1.1 sgRNAs9 软件图形用户界面编制和运行环境

sgRNAs9 软件包为 Perl 语言(<https://www.perl.org/>)编写。基于 NetBeans IDE 8.0.2 程序开发平台(<https://netbeans.org/>),针对其核心运行程序“sgRNAs9.pl”,采用 Java 程序语言(https://java.com/zh_CN/)编写了图形用户界面,同时还对其脱靶风险评估算法和结果输出源代码进行了优化。除细菌和病毒基因组较小外,一般物种的全基因组序列大小约为 2~4 Gb,软件运行时会将基因组序列读入内存。因此使用 sgRNAs9 软件时,建议用户的计算机内存应不小于 4 GB,否则可能会因“内存不足”而导致软件无法

正常运行。该软件可在 Windows 不同版本操作系统上运行,用户需提前安装 Perl 和 Java 程序。

1.2 sgRNA 设计和脱靶效应评估

为设计靶向 miRNA 前体的 sgRNA 并分析其特征,从 miRNA 数据库 miRBase V21.0 (<http://www.mirbase.org/>)下载所有物种的 miRNA 前体序列^[25],利用自编的 Perl 程序分别提取人、小鼠、大鼠、猪和鸡等 5 个物种所有 miRNA。从 Ensembl 数据库^[34]下载人的全基因组序列(ftp://ftp.ensembl.org/pub/release-80/fasta/homo_sapiens/dna/)。然后使用 sgRNAs9 软件设计 sgRNA,大致操作流程为:输入含有上述 5 个物种 miRNA 前体序列的 FASTA 格式文件路径,进行参数设置,如选择长度为 20nt 的 sgRNA,GC%含量为 20%~80%,另外选择单个 sgRNA(Single sgRNA)和 DNA 双链(Both DNA strand)设计模式。由此,可针对 miRNA 前体的正负链,设计长度为 20nt,且 GC%为 20%~80%的 sgRNA。最后对靶向人 hsa-mir-206 前体的 sgRNA 进行脱靶效应评估。采用该软件默认的脱靶评估参数,即脱靶位点最大允许含 5 个错配碱基,PAM 类型为“NGG”或“NAG”,同时评估 sgRNA 种子序列的特异性,并自动排除含“TTTT”终止信号的 sgRNA。

1.3 引物设计与载体构建

为验证靶向人 hsa-mir-206 前体 sgRNA 的基因组编辑活性和脱靶效应,利用 sgRNAs9 软件包(V3.0.5)中的 sgRNA 引物设计程序“sgRPrimer_v2.0.pl”,针对 pGL3-U6-sgRNA 载体,设计含 *Bsa* I 限制性内切酶(New England Biolabs)识别位点的引物以构建 sgRNA 过表达载体(表 1),其中 pGL3-U6-sgRNA 质粒由上海科技大学的黄行许教授实验室馈赠^[35]。使用靶标位点侧翼序列提取程序“extract_targetSeq.pl”,批量提取 sgRNA 打靶和脱靶位点两侧各 500~700 bp 的基因组序列,再用 Primer Premier 5.0 软件设计最佳的 PCR 扩增引物对(表 1),将引物送北京天一辉远生物科技有限公司合成。sgRNA 表达载体构建大致流程为:取浓度为 100 $\mu\text{mol/L}$ 的“hsa-mir-206_S_4-F”和“hsa-mir-206_S_4-R”引物各 1 μL ,加 8 μL H_2O ,在 PCR 仪上用 95 $^{\circ}\text{C}$ 加热 5 min,然后室温放置 1 h 冷却。取 2 μL 退火的引物对与提前用 *Bsa* I 酶切的 pGL3-U6-sgRNA 质粒,使用 T4 DNA

表 1 靶向人 hsa-mir-206 前体 sgRNA 载体构建和打靶及脱靶位点检测引物

引物名称	引物序列(5'→3')	扩增长度(bp)
hsa-mir-206_S_4-F	accgATGGAATGTAAGGAAGTGTG	
hsa-mir-206_S_4-R	aaacCACACTTCCTTACATTCCAT	
On-target-F	GAGTGGCTCTCTGCGTGAAT	602
On-target-R	GCTTCCTTGGTGAGGGAGTC	
OT-1-F	TGGCAGTGAGGGAACATGAC	785
OT-1-R	GGCACTTGGTGCTGGAATG	
OT-2-F	TGTCACCCGACACAATGAGG	783
OT-2-R	AGTGAGTTAGTGACACCCA	
OT-3-F	GATCAGTGAGGGTTGGGCAT	643
OT-3-R	TCCAGCCCCAACTTGAATC	
OT-4-F	GCAGTAAAGGCAAGTGCTGT	701
OT-4-R	TGTGCTGTCTGATTGAGACC	
OT-5-F	AGGTCACCAGAGATCCCTTC	759
OT-5-R	CACTTTTCTGCTCGGTCCCT	
OT-6-F	ACCTGTCTTGAAGCTCAAGG	799
OT-6-R	GGCCCTACATGTGGTGAAAG	
OT-7-F	CGCTGTGACTGAGGATGGAG	701
OT-7-R	GCAAAAGGAAGCCATGGGAC	
OT-8-F	ATCTGGGGAGTGAGGATGGT	842
OT-8-R	GAGCCATTCAGGCCAGCTTT	
OT-9-F	CTGAGCTTCAGACACCTCTCG	800
OT-9-R	GCGTCTCACATTTGCACTCA	
OT-10-F	TCCTCTTATTGCGCCAAGGT	859
OT-10-R	CCAACTCCCCACTTTACCCA	
OT-11-F	GAGCCTCTGGAGGCAGTAATC	701
OT-11-R	GCACAGGCCTAGCATCTTTG	
OT-12-F	CTGGCAGCTTGAAATTGGT	797
OT-12-R	AACCTTCAGGGGATTGTGCT	
OT-13-F	CCTGAACCCAATTCCACACTC	932
OT-13-R	CCTCCTAGTCTCTGCACTC	
U6seqF	ACTATCATATGCTTACCGTAAC	

注：hsa-mir-206_S_4-F/R 为构建 sgRNA 表达载体引物，小写字母表示 *Bsa*I 酶切识别位点；On-target-F/R 为打靶位点检测引物；OT-F/R 为脱靶位点检测引物；U6seqF 为测序引物。

连接酶(宝生物工程(大连)有限公司)16℃连接过夜。取 1~2 μL 连接产物，转化到 DH5α 感受态细胞中，挑选用 PCR 鉴定成功的单克隆菌液，送生工生物工程(上海)股份有限公司用“U6seqF”引物测序验证(表 1)，新构建的载体命名为“pGL3-U6-sgRNA-hsa-mir-206_S_4”。最后按照去内毒素质粒大抽试剂盒(Omega Bio-Tek)说明书，分别抽提 sgRNA 和 Cas9 核酸酶表

达质粒，并测定其浓度。其中，单独表达 Cas9 核酸酶的质粒 pST1374-Cas9-ZF-NLS 同样由上海科技大学的黄行许教授实验室馈赠^[35]。

1.4 HEK 293T 细胞培养和转染

HEK 293T 细胞购自中国科学院细胞库。用含 10% 胎牛血清的 DMEM(Invitrogen Life Science)培养基，在 CO₂ 浓度为 5%，温度为 37℃的培养箱中培养。按照 Lipofectamine[®]2000 转染试剂(Invitrogen Life Science)说明书进行分组转染，第 1 组转染 pST1374-Cas9-ZF-NLS 质粒，仅表达 Cas9 核酸酶；第 2 组按等质量比转染 pST1374-Cas9-ZF-NLS 和 pGL3-U6-sgRNA-hsa-mir-206_S_4，同时表达 sgRNA 和 Cas9 核酸酶。转染 48 h 后收集细胞，利用基因组 DNA 小量抽提试剂盒(天根生化科技(北京)有限公司)，抽提不同实验组和未做处理的细胞基因组 DNA。

1.5 PCR 扩增与 T7E1 酶切实验

利用表 1 中检测打靶和脱靶的引物对，用上述提取的 DNA 为模板，利用高保真 *Taq* DNA 聚合酶(宝生物工程(大连)有限公司)进行 PCR 扩增，PCR 反应条件为：95℃预变性 5 min，94℃变性 30 s，60℃复性 30 s，72℃延伸 30 s，扩增 30 个循环。利用 PCR 纯化试剂盒(天根生化科技(北京)有限公司)纯化 PCR 产物，补充一定体积的 NEBuffer 2 溶液(New England Biolabs)后，在 PCR 仪上进行梯度复性杂交，程序为：95℃变性 5 min；95–85℃，每秒降低 2℃；85–25℃，每秒降低 0.1℃；4℃保存^[34]。取 0.5 μL T7E1(New England Biolabs)限制性内切酶，37℃酶切杂交的 PCR 产物 15 min，然后在 2%的琼脂糖凝胶电泳检测，突变的泳道会出现小分子量的条带。进一步，利用 TA 克隆方法将利用打靶引物对(On-target-F/R)扩增的 PCR 产物，连接进 pMD18-T 载体(宝生物工程(大连)有限公司)中，转化大肠杆菌细胞，挑选单克隆在 LB 培养基中扩大培养，对培养的菌液进行 PCR 鉴定后送生工生物工程(上海)股份有限公司测序。

2 结果与分析

2.1 sgRNACas9 软件图形用户界面开发

sgRNACas9 软件包共有 8 个功能不同且可独立

运行的小程序,均为 Perl 语言编写。为了方便用户使用 sgRNAs9 软件设计 sgRNA 并进行脱靶效应评估,基于 NetBeans IDE 8.0.2 程序开发平台,针对核心程序“sgRNAs9.pl”,本文采用 Java 语言开发了可在微软 Windows 操作系统运行的图形用户界面。由图 1 可见,该图形用户界面主要由三部分组成:输入(Input),参数设置(Parameter settings)和运行控制(RUN 和 EXIT)。其中“输入”窗口含两个框:靶标序列(Target sequence)和基因组序列(Genome sequence),均需输入对应 FASTA 格式序列文件的完整路径。用户也可直接点击“选择(choose)”按钮确定目标文件。值得一提的是 FASTA 指 DNA 序列的第一行开始于一个标识符:“>”,紧接着(没有空格)是对该序列的唯一描述(即 ID),从第二行开始就是序列。

2.2 图形用户界面参数设置及说明

对于“参数设置”窗口(图 1),共含有 6 个可供用户设置参数的选项,分别为:(1)选择 sgRNA 长度(Select sgRNA length (17~20nt)),此处可设置 sgRNA 的长度,可选范围为 17~20nt,默认参数值为 20nt。

(2)GC 含量百分比(GC%),用户可设置最小(MIN)和最大值(MAX),如图 1 所示,默认参数值为 40%~60%。(3)选择合适的 DNA 链(Select appropriate DNA strand(s)),对应可选的项目有:A,DNA 正义链(Senses DNA strand);B,DNA 反义链(Anti-senses DNA strand);C,DNA 双链(both DNA strands)。值得一提的是,此处的正义链和反义链不具有生物学意义,仅针对用户输入的靶基因序列的方向,正义链表示序列方向为 5'→3',而反义链表示其反向互补链,软件会自动转换输入的序列。(4)sgRNA 类型(Type of gRNA),对应可选的项目有:A,单个 sgRNA(single sgRNA);B,Paired-gRNAs,其中 paired-gRNA 主要针对 Cas9 单切口酶^[35]和融合有 *Fok I* 的 Cas9 核酸酶^[36],这两种类型的 Cas9 核酸酶对两个 sgRNA 在 DNA 链上的位置和距离均有一定的要求,可通过(5)paired-gRNA 的间距(Offset distance of paired-gRNAs)选项进行设置。如图 1 所示,针对 Cas9 单切口酶,软件默认两个位于不同 DNA 单链上的 sgRNA 间距为-2~32 bp。如果针对融合有 *Fok I* 的 Cas9 核酸酶,已有研究表明^[36],当 paired-gRNA 的

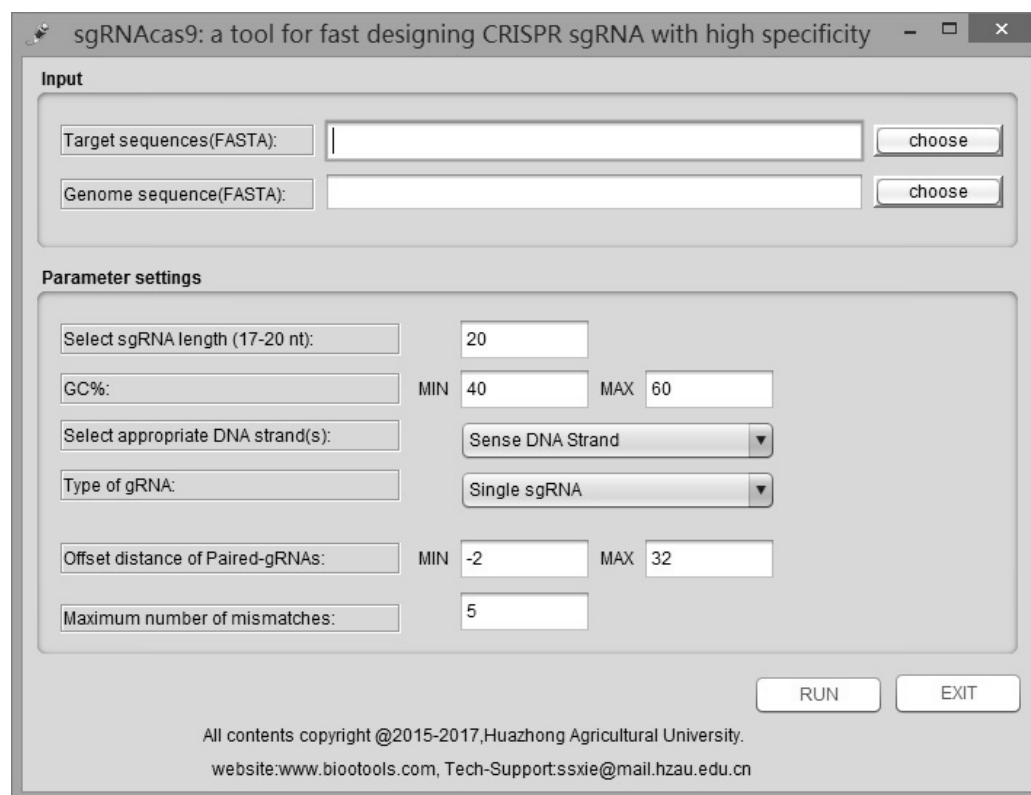


图 1 sgRNAs9 软件的图形用户界面

距离接近 16 bp 时,其基因组编辑效率最大,因此用户可将此参数设置为 15~17 bp。值得指出的是,对于是否设置“paired-gRNA 的间距”选项,主要取决于用户是否选择将“gRNA 类型”选项中的参数设置为“Paired-gRNAs”。(6)最大碱基错配数(Maximum number of mismatches)选项,用于设置脱靶位点最大允许的错配碱基数,默认参数值为 5。

2.3 sgRNA 脱靶风险分类与软件运行结果优化

依据脱靶位点碱基错配数和位置以及种子序列的特异性,本文将 sgRNA 的脱靶风险划分了 6 个不同等级(表 2),分别是“弃用(Discard)”、“高风险(High_risk)”、“中等风险(Moderate_risk)”、“低风险(Low_risk)”、“识别位点重复或较差(Repeat_sites_or_bad)”和“最佳(Best)”。值得一提的是,对于软件评估为“弃用(Discard)”的 sgRNA,表明该 sgRNA 在基因组上无结合位点。出现此情况可能的原因是用户使用的靶基因序列为 mRNA,由此设计的 sgRNA 可能位于两个外显子的连接处,即跨越内含子而使其无法在基因组上找到对应的靶标位点,应予以舍弃。而对于软件评估为“识别位点重复或较差(Repeat_sites_or_bad)”的 sgRNA,表明该 sgRNA 可同时靶向多个位点。是否选择该 sgRNA 主要基于以下两点:(1)如果目标基因是通过基因重复产生的两个或更多的拷贝,针对该多拷贝基因设计的 sgRNA 作为候选进行多拷贝基因敲除。(2)如果该 sgRNA 靶向位于不同基因组位置的多个完全匹配的位点,且对应不相关的基因或位点,则其脱靶风险最大,此类 sgRNA 将予以舍弃。此外,软件会根据脱靶位点碱基错配数量和种子序列的特异性,进一步将 sgRNA 的脱靶风险划分不同等级(表 2)。

此外,最初的 sgRNAs9 核心程序(V2.0.6)输出文件数量较多。为方便用户使用 sgRNAs9 软件,我们对其核心程序进行了修改,对结果输出文件进行了整合,将文件数从原有的 20 个减少至 9 个,分别是“A.Sort_OT_byID”,“B.Sort_POT_byID”,“POT_seed.identity”,“CRISPR.targets_single.fa”,“report_protospacer_pairs.xls”,“report_protospacer_single.txt”,“sgRNAs9_report.xls”,“TargetSeq.fa”。值得一提的是,修改后的软件将 sgRNA 序列信息和脱靶效应评估数据集中到“sgRNAs9.report.xls”一个表格中,可

表 2 sgRNA 脱靶风险按等级高低分类

sgRNA 脱靶风险	分类依据
弃用(Discard)	在基因组上无对应的靶标结合位点
高风险(High_risk)	存在 1 个碱基错配的脱靶位点,其中含种子序列完全匹配,且非种子序列仅含 1 个碱基错配的脱靶位点
中等风险(Moderate_risk)	存在 2 个碱基错配的脱靶位点,其中含种子序列完全匹配,且非种子序列含不多于 2 个碱基错配的脱靶位点
低风险(Low_risk)	存在多于 2 个碱基错配的脱靶位点
识别位点重复或较差(Repeat_sites_or_bad)	存在多个完全匹配的靶标结合位点
最佳(Best)	不存在少于 2 个碱基错配的脱靶位点,且其中不含种子序列完全匹配的脱靶结合位点

方便用户仅需依据脱靶风险等级和碱基错配总数两个指标挑选特异性高的 sgRNA。sgRNAs9 图形用户版软件运行完成后,结果将输出到“report”文件夹中。如果软件运行出现异常而无结果输出,则可在“logs”文件夹中查看“sgRNAs9.log”日志文件,其记录有 sgRNAs9 软件的运行状态,由此可判定软件运行时何处出错。

2.4 靶向 5 个物种 miRNA 前体 sgRNA 的特征

利用 sgRNAs9 软件,本文针对 miRBase 数据库 V21(<http://www.mirbase.org/>)所收录的人(1881)、小鼠(1193)、大鼠(495)、猪(382)和鸡(740)等 5 个物种的 miRNA 前体,分别设计了 13 355、7957、3931、3014 和 5867 条相应的 sgRNA(附表 1)。统计分析发现,并非所有 miRNA 前体均含有 sgRNA 靶标位点。在附表 1 中,不同物种含有 sgRNA 靶标位点的 miRNA 前体数目依次是:人(1798)、小鼠(1115)、大鼠(486)、猪(376)和鸡(725)。换言之,人有 83 个、小鼠有 78 个、大鼠有 9 个、猪有 6 个和鸡有 15 个 miRNA 前体无 sgRNA 靶标位点。此外,对所有 34124 条 sgRNA 的 GC%含量进行分析,结果发现设计的 sgRNA 的 GC%范围为 20%~75%,且相对集中于 40%~60%(图 2A)。其中 GC%含量为 55%的 sgRNA 占比最高为 15.3%;其次是 GC%含量为 60%的 sgRNA,其比例为 14.9%,而 GC%含量为 50%的占总比例为 14.3%(图 2A)。进一步对单个 miRNA 前体所含有的 sgRNA 靶标位点个数进行统计分析,结果发现单个 miRNA 前体含 7 个 sgRNA 靶标位点的比

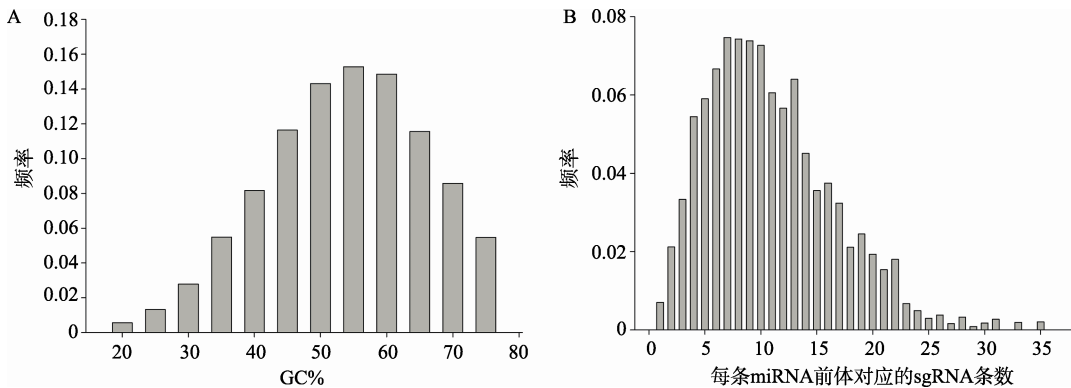


图 2 靶向 4691 个 miRNA 前体的 sgRNA 特征
A : 34 124 条靶向不同物种共 4691 个 miRNA 前体 sgRNA 的 GC%含量分布 ; B : 单个 miRNA 前体含有的 sgRNA 靶标位点数分布。

例最高, 所占比例为 7.5%, 而含 8 个 sgRNA 靶标位点的比例次之(7.4%)。其中 0.7%的 miRNA 前体仅含有 1 个 sgRNA 靶标位点, 0.2%的 miRNA 前体含多达 35 个 sgRNA 靶标位点(图 2B)。

2.5 靶向人 miR-206 前体 sgRNA 的基因组编辑活性验证

为验证 sgRNAs9 软件设计的 sgRNA 是否具有基因组编辑活性, 针对人的 hsa-mir-206 前体(86 bp), 设计了 11 条 sgRNA 并进行脱靶效应评估(附表 2), 并挑选了针对 miR-206 成熟序列编号为“mir-206_S_4”的 sgRNA 进行打靶和脱靶活性验证(图 3A)。T7E1 酶切实验表明, 该 sgRNA 能引导 Cas9 核酸酶切割 hsa-mir-206 基因组序列(图 3B)。随后, 将 PCR 产物连接到 pMD18-T 载体中, 随机挑选了

20 个菌液送生工生物工程(上海)股份有限公司进行测序。测序结果显示其中有 6 个测序样本序列存在碱基改变。如图 4 所示, 编号为“1#”的序列在 sgRNA 的靶标位点附近发生了 1 个碱基突变, 同时缺失了 10 个碱基(图 4A)。编号为“2#、3#、4#和 5#”的序列均发现有碱基突变, 而编号为“6#”的序列发生了单碱基插入(图 4A)。

随后利用 mfold 软件^[37](<http://mfold.rna.albany.edu/?q=mfold>), 分别对碱基缺失(1#)和单碱基插入(6#)的 hsa-mir-206 前体序列进行了二级结构折叠(图 4B), “1#”因缺失的 10 个碱基正好位于 miR-206 成熟序列中, 这不仅破坏了成熟 miRNA 序列, 而且使其茎环结构发生改变, 使得 Gibbs 自由能(ΔG)由 -43.3 kcal/mol 变为 -24.6 kcal/mol , 推测此突变很可能影响 miR-206 的加工成熟及其功能。尽管“6#”序

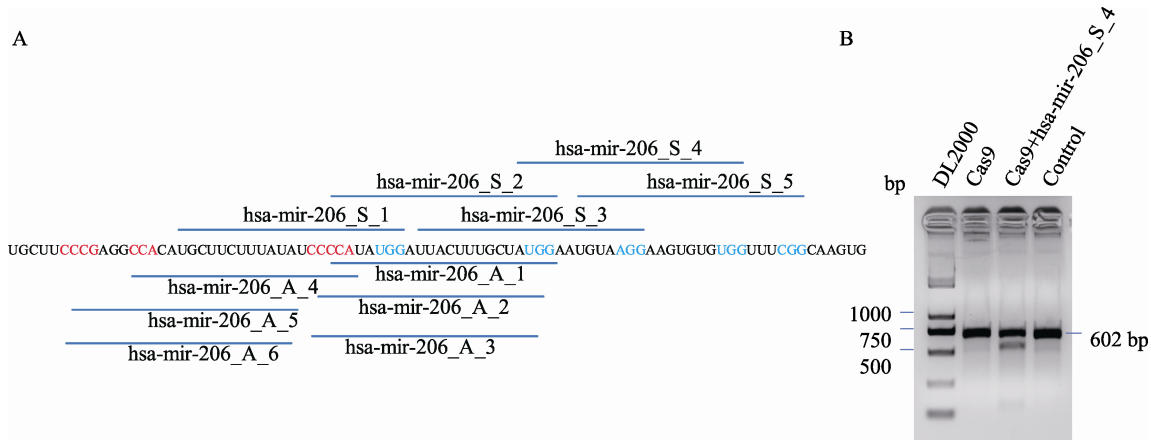


图 3 靶向人 hsa-mir-206 前体 sgRNA 的位置及其活性验证
A : 靶向 hsa-mir-206 前体的 sgRNA 分布示意图 ; B : T7E1 酶切实验。DL2000 , DNA marker ; Cas9 , 表示 pST1374-Cas9-ZF-NLS ; Cas9+ hsa-mir-206_S_4 , 表示 pST1374-Cas9-ZF-NLS 和 pGL3-U6-sgRNA-hsa-mir-206_S_4 ; Control , 表示未转染载体的阴性对照。

A

CCCATATGGATTACTTTGCTATGGAATGTAAGGAAGTGTGTTTCGGCAAGTGCCTCCTCGCTGCCCCAGGGTACCA (Wild type)

1# CCCATATGGACTACTTTGCTATGGAA-----TGTGTGGTTTCGGCAAGTGCCTCCTCGCTGCCCCAGGGTACCA (1 base mutation and 10 base deletion, 1/20)

2# CCCATATGGATCACTTTGCTATGGAATGTAAGGAAGTGTGTGGTTTCGGCAAGTGCCTCCTCGCTGCCCCAGGGTACCA (1 base mutation, 1/20)

3# CCCATATGGATTACTTTGCTATGGAATGTAAGGAAGTGTGTGGTTTCGGCAAGTGCCTCCTCGCTAGCCCCAGGGTACCA (1 base mutation, 1/20)

4# CCCATATGGATTACTTTGCTATGGAATGTAAGGAAGTGTGTGGTTTCGGCAAGTGCCTCCTCGCTGGTCCAGGGTACCA (1 base mutation, 1/20)

5# CCCATATGGATTACTTTGCTATGGGATGTAAGGAAGTGTGTGGTTTCGGCAAGTGCCTCCTCGCTGCCCCAGGGTACCA (1 base mutation, 1/20)

6# CCCATATGGATTACTTTGCTATGGAATGTAAGGAAG\T\TGTGTGGTTTCGGCAAGTGCCTCCTCGCTGCCCCAGGGTACCA (1 base insertion, 1/20)

B

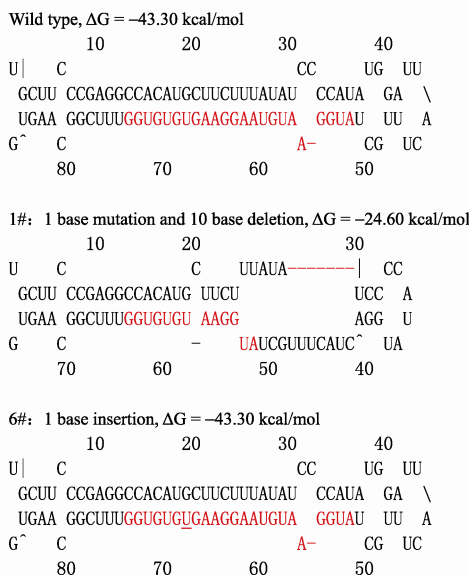


图 4 人细胞中 hsa-mir-206 前体不同类型碱基改变及对二级结构折叠的影响

A: 6 种类型 hsa-mir-206 基因型; B: miRNA 前体二级结构折叠。其中前体序列变异类型有碱基突变(mutation)、缺失(deletion)和插入(insertion), Wild type, 表示野生型, 1/20 表示随机挑取 20 个克隆进行测序, 出现某种基因型的概率为 1。

列的单碱基插入到 miR-206 成熟序列中, 但折叠后发现其茎环结构未受影响, 推测插入的碱基可能不会影响 miR-206 的加工成熟, 但可能会改变其生物学功能。另外“5#”序列突变也正好位于 miR-206 的成熟序列, 其也可能改变该 miRNA 的功能。而编号为“2#、3#和 4#”的序列突变碱基均位于 hsa-mir-206 前体中, 可能仅影响该 miRNA 的加工成熟, 但具体不同类型的碱基改变会造成何影响, 还须进一步通过实验验证。总之, 以上实验数据表明 sgRNAs9 软件设计的 sgRNA 具有打靶活性。

2.6 靶向人 miR-206 前体 sgRNA 的脱靶位点预测及验证

利用 sgRNAs9 软件, 对上述编号为“hsa-mir-206_S_4”的 sgRNA 脱靶效应进行了预测。由附表 2 可见, 该 sgRNA 可预测到含 5 个以内碱基错配的脱

靶位点总数有 1237 条, 其中有 2 个脱靶位点仅含 2 个碱基错配, 22 个脱靶位点含 3 个碱基错配, 376 个脱靶位点含 4 个碱基错配, 而有 837 个脱靶位点含 5 个碱基错配。对该 sgRNA 种子序列的特异性进行分析, 发现含种子序列一致的脱靶位点共有 15 个, 其中包含 1 条有 3 个碱基错配的脱靶位点, 12 条有 4 个碱基错配的脱靶位点和 2 条有 5 个碱基错配的脱靶位点。由附表 2 的风险评估显示, 该 sgRNA 的脱靶风险等级为中等风险(Moderate_risk)。附表 3 标示了该 sgRNA 预测的脱靶位点在基因组上的位置, 以及每个脱靶位点碱基错配数量及其错配位置。

为对上述预测的脱靶位点进行验证, 本文随机挑选了 13 个脱靶位点(表 3), 先分别通过 PCR 对这些位点进行扩增, 再利用 T7E1 限制性内切酶进行酶切实验。结果发现, 其中有 2 个预测的脱靶位点, 即编号为“OT-9”和“OT-12”的位点可被 T7E1 酶切实

表 3 靶向人 miR-206 前体 sgRNA 部分预测的脱靶位点及 T7E1 酶切验证

has-mir-206	序列(5'→3')	碱基错配数	染色体	位置	DNA 链	T7E1 实验
On target	ATGGAATGTAAGGAAGTGTG-TGG		6	52144349-52144434	+	Yes
OT-1	AAGGAACTAAGGAAGTGTG-GGG	3	14	75929214	+	No
OT-2	GTGGTACTAAGGAAGTGTG-TGG	5	1	226764668	-	No
OT-3	CGGGAAGTTAAGGAAGTGTG-GGG	4	3	149989212	+	No
OT-4	AAATATTGTAAGGAAGTGTG-TGG	4	4	72188804	+	No
OT-5	ATAAGTCTAAGGAAGTGTG-GGG	4	8	41007917	-	No
OT-6	ATAGCATGGAGGAAGTGTG-TGG	4	10	35268699	+	No
OT-7	ATGAAAAAAGGAAGTGTG-AGG	4	11	60713450	-	No
OT-8	ATGGTCTGTAAATAAGTGTG-TAG	4	10	97968951	-	No
OT-9	CTGGCATGTACGAAGTGTG-GGG	4	12	15166238	-	Yes
OT-10	ATTATTGGGAAGTGTG-TGG	5	10	113140506	+	No
OT-11	ATGTGATGACAGAAGTGTG-TGG	5	11	119230605	+	No
OT-12	ATGGAATGGCAGGAAGTGTG-TGG	3	12	15347488	-	Yes
OT-13	ATGGCAGGTAAGGAAGTGAAG-GGG	3	13	28781365	-	No

注：OT，即 off-target site，表示为预测的脱靶位点。序列中的红色标记代表错配碱基。

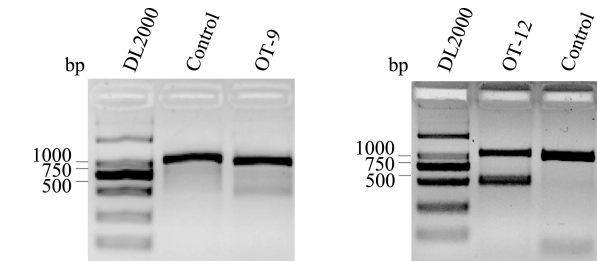


图 5 T7E1 酶切实验检测预测的脱靶位点
OT，脱靶位点(off-target site)；DL2000，DNA marker；Control，表示未转染载体的阴性对照。

验所证实(图 5)，表明这两个位点发生了脱靶，且“OT-12”的脱靶程度相对较高。由表 3 可见，脱靶位点“OT-9”有 4 个碱基错配，而“OT-12”仅有 3 个碱基错配。

3 讨 论

基于 CRISPR/Cas9 系统介导的基因组编辑技术的开发与应用，已经使动、植物和微生物等诸多物种的基因工程操作手段发生了革命性的改变。尽管 CRISPR/Cas9 技术强大，但因其存在脱靶效应，制约了其进一步的应用与发展。因此，如何提高 CRISPR/Cas9 系统介导的基因组编辑效率和特异性是本领域的研究热点。本研究针对 sgRNAs9 软件包中的核心运行程序，开发了其图形用户界面。此外，还对 sgRNAs9 软件的核心程序进行了修改。

新修改的程序对于脱靶效应评估综合考查两个因素：

(1)脱靶位点的碱基错配数量，最大允许 5 个碱基错配；(2)靠近 3'端临近 PAM 长度为 12 bp 种子序列的特异性。基于此，在 sgRNAs9 图形用户界面版软件中，新引入了一个包括 6 个等级的风险评估指标(表 2)。用户可根据对应的风险等级选择合适的 sgRNA，建议选择顺序依次是：最佳(Best)>低风险(Low_risk)>中等风险(Moderate_risk)>高风险(High_risk)>识别位点重复或较差(Repeat_sites_or_bad)>弃用(Discard)。其中，对于评定为“识别位点重复或较差(Repeat_sites_or_bad)”的 sgRNA，用户需要进一步确定该 sgRNA 的靶点是否为多拷贝基因，以决定是否选择或弃用。值得注意的是，假如有多个评定为“最佳(Best)”的 sgRNA 供选择，用户可再通过排序选择总脱靶数量最少的 sgRNA 用于实验。另外，还可依据 sgRNA 在基因上的位置，选择满足实验需求的 sgRNA，如为了造成蛋白编码基因的移码突变，可在 mRNA 蛋白编码区起始密码子(ATG)下游约 100~200 bp 位置设计 sgRNA。

miRNA 是一类小分子非编码调控性 RNA，其参与机体内各种生物学进程^[38,39]。已有报道指出，利用 CRISPR/Cas9 技术，可以破坏 miRNA 的前体或成熟体，抑制其加工成熟和功能^[40,41]。国内已有学者应用 CRISPR/Cas9 基因编辑技术，成功建立了 miRNA-

301a 敲除小鼠^[42]。鉴于 miRNA 前体序列范围是 50~180 bp, 为探明较短长度的 miRNA 前体序列上是否可寻找到 sgRNA 靶标位点, 本文利用 sgRNAs9 软件, 对从 miRBase 数据库来源的人、小鼠、大鼠、猪和鸡的共 4691 个 miRNA 前体设计 sgRNA。分析发现, 大部分 miRNA 前体均含有“N₂₀NGG”的靶标识别位点, 但所有 5 个物种中, 均有少数 miRNA 前体无 sgRNA 靶标位点, 如 hsa-mir-8485, hsa-mir-718 和 hsa-mir-6870 等。理论上讲, 针对 miRNA 前体, 乃至成熟序列设计靶向 sgRNA, 可达到干扰 miRNA 加工成熟的目的。而对于无法在前体中找到 sgRNA 靶标位点的 miRNA, 可采用在前体两侧设计 sgRNA, 由此可将整个 miRNA 前体对应的基因组序列切除。另外, 对设计靶向 miRNA 前体的 sgRNA 的 GC% 含量进行分析发现, 其范围集中于 55%。已有研究表明^[43], 活性较高 sgRNA 的 GC% 范围为 40%~60%。由此可见, 设计的大部分 sgRNA 的 GC% 含量位于适用范围, 推测这些 sgRNA 具有活性。此外, 还发现大多数 miRNA 前体含 7~8 个 sgRNA 靶标位点, 也有少数仅含 1 个靶标位点, 而有些竟含多达 35 个靶标位点。而理论上讲, 含多个 sgRNA 靶标位点的 miRNA, 更有利于寻找到脱靶风险低的 sgRNA。进一步分析发现, miRNA 前体含 sgRNA 靶标位点数量多少, 不仅与其自身序列有关, 而且与前体序列的长度有关。比如, 含 35 个 sgRNA 靶标位点的 hsa-mir-6753 长度为 164 bp, 而含 1 个靶标位点的 hsa-mir-18a 长度为 71 bp。尽管本文针对人和小鼠等 5 个物种 miRNA 前体设计了 34124 条 sgRNA, 但鉴于计算机硬件条件限制, 并未继续对这些 sgRNA 进行全基因组水平脱靶效应评估。

为检测 sgRNAs9 软件设计的 sgRNA 是否具有活性并验证软件预测的脱靶位点是否真实存在, 本文随机挑选了 1 个靶向 hsa-miR-206 成熟体的 sgRNA 在细胞水平进行了验证。结果表明, 该 sgRNA 具有基因组编辑活性。通过测序发现, 有一条 miR-206 序列的 miRNA 成熟体存在碱基缺失, 推测该 miRNA 的功能会被破坏。但有的 miRNA 序列仅发生碱基突变, 且突变位点位于 miRNA 前体中, 其是否能抑制 miRNA 的转录与加工成熟需进一步验证。对该 sgRNA 进行脱靶效应预测, 发现其全基因组水平含小于或等于 5 个碱基错配的脱靶位点数量大于 1000

个。对随机选择的 13 个脱靶位点进行验证, 其中包含有种子序列一致的脱靶位点(表 3), 结果发现有 2 个脱靶位点(分别含 3 和 4 个碱基错配)可被实验验证。而种子序列一致的 6 个预测脱靶位点未发生脱靶(表 3)。理论上讲, 含种子序列一致的脱靶位点极容易发生脱靶, 但我们的实验结果与此不符。由此可见, 尽管 sgRNAs9 软件可预测到真实的脱靶位点, 但就验证数据来看, 准确率并不高。推测可能的原因: 一方面, 本文验证的脱靶位点数量有限, 需加大实验样本量; 其次, 目前对脱靶效应的分子机理尚不清楚, 需要进一步通过实验研究, 以获得更多数据, 进而改进软件的脱靶效应评估算法并提高预测的准确率。总之, sgRNAs9 图形用户界面软件的开发, 必将有助于广大科研工作者使用该软件设计特异性的 sgRNA。

参考文献

- [1] Brouns SJJ, Jore MM, Lundgren M, Westra ER, Slijkhuys RJH, Snijders APL, Dickman MJ, Makarova KS, Koonin EV, van der Oost J. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science*, 2008, 321(5891): 960–964. [DOI]
- [2] 李君, 张毅, 陈坤玲, 单奇伟, 王延鹏, 梁振, 高彩霞. CRISPR/Cas 系统: RNA 靶向的基因组定向编辑新技术. *遗传*, 2013, 35(11): 1265–1273. [DOI]
- [3] Bibikova M, Golic M, Golic KG, Carroll D. Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics*, 2002, 161(3): 1169–1175. [DOI]
- [4] Li T, Huang S, Zhao XF, Wright DA, Carpenter S, Spalding MH, Weeks DP, Yang B. Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res*, 2011, 39(14): 6315–6325. [DOI]
- [5] 沈延, 肖安, 黄鹏, 王唯晔, 朱作言, 张博. 类转录激活因子效应物核酸酶(TALEN)介导的基因组定点修饰技术. *遗传*, 2013, 35(4): 395–409. [DOI]
- [6] 沈延, 黄鹏, 张博. TALEN 构建与斑马鱼基因组定点突变的实验方法与流程. *遗传*, 2013, 35(4): 533–544. [DOI]
- [7] 唐成程, 张全军, 李小平, 樊娜娜, 杨翌, 全龙泉, 赖良学. 应用 TALEN 技术对兔 CCR5 基因进行靶向修饰. *遗传*, 2014, 36(4): 360–368. [DOI]
- [8] 李飞达, 李勇, 刘欢, 张欢欢, 刘楚新, 张兴举, 窦红伟, 杨文献, 杜玉涛. 利用 TALENs 和手工克隆技术高效获

- 得 GHR 基因敲除巴马猪. 遗传, 2014, 36(9): 903–911. [DOI]
- [9] 杨翠翠, 佟慧丽, 马兴红, 杜巍, 刘丹, 杨宇, 严云勤. 利用 TALEN 技术在牛胎儿成纤维细胞中敲除 *Myostatin* 基因. 遗传, 2014, 36(7): 685–690. [DOI]
- [10] Niu YY, Shen B, Cui YQ, Chen YC, Wang JY, Wang L, Kang Y, Zhao XY, Si W, Li W, Xiang AP, Zhou JK, Guo XJ, Bi Y, Si CY, Hu B, Dong GY, Wang H, Zhou ZM, Li TQ, Tan T, Pu XQ, Wang F, Ji SH, Zhou Q, Huang XX, Ji WZ, Sha JH. Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. *Cell*, 2014, 156(4): 836–843. [DOI]
- [11] Zhou JK, Shen B, Zhang WS, Wang JY, Yang J, Chen L, Zhang N, Zhu K, Xu J, Hu B, Leng QB, Huang XX. One-step generation of different immunodeficient mice with multiple gene modifications by CRISPR/Cas9 mediated genome engineering. *Int J Biochem Cell Biol*, 2014, 46: 49–55. [DOI]
- [12] Ma YW, Zhang X, Shen B, Lu YD, Chen W, Ma J, Bai L, Huang XX, Zhang LF. Generating rats with conditional alleles using CRISPR/Cas9. *Cell Res*, 2014, 24(1): 122–125. [DOI]
- [13] Hai T, Teng F, Guo RF, Li W, Zhou Q. One-step generation of knockout pigs by zygote injection of CRISPR/Cas system. *Cell Res*, 2014, 24(3): 372–375. [DOI]
- [14] Chang NN, Sun CH, Gao L, Zhu D, Xu XF, Zhu XJ, Xiong JW, Xi JJ. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. *Cell Res*, 2013, 23(4): 465–472. [DOI]
- [15] Jiang WZ, Zhou HB, Bi HH, Fromm M, Yang B, Weeks DP. Demonstration of CRISPR/Cas9/sgRNA-mediated targeted gene modification in Arabidopsis, tobacco, sorghum and rice. *Nucleic Acids Res*, 2013, 41(20): e188. [DOI]
- [16] 王延鹏, 程曦, 高彩霞, 邱金龙. 利用基因组编辑技术创制抗白粉病小麦. 遗传, 2014, 36(8): 848. [DOI]
- [17] Ebina H, Misawa N, Kanemura Y, Koyanagi Y. Harnessing the CRISPR/Cas9 system to disrupt latent HIV-1 provirus. *Sci Rep*, 2013, 3: 2510. [DOI]
- [18] Lin SR, Yang HC, Kuo YT, Liu CJ, Yang TY, Sung KC, Lin YY, Wang HY, Wang CC, Shen YC, Wu FY, Kao JH, Chen DS, Chen PJ. The CRISPR/Cas9 system tates clearance of the intrahepatic HBV templates in vivo. *Mol Ther Nucleic Acids*, 2014, 3: e186. [DOI]
- [19] 殷利眷, 胡斯奇, 郭斐. CRISPR-Cas9 基因编辑技术在病毒感染疾病治疗中的应用. 遗传, 2015, 37(5): 412–418. [DOI]
- [20] Gilbert LA, Larson MH, Morsut L, Liu ZR, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 2013, 154(2): 442–451. [DOI]
- [21] Maeder ML, Linder SJ, Cascio VM, Fu YF, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. *Nat Methods*, 2013, 10(10): 977–979. [DOI]
- [22] Fu YF, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol*, 2013, 31(9): 822–826. [DOI]
- [23] Wu XB, Kriz AJ, Sharp PA. Target specificity of the CRISPR-Cas9 system. *Quant Biol*, 2014, 2(2): 59–70. [DOI]
- [24] Pattanayak V, Lin S, Guilinger JP, Ma EB, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol*, 2013, 31(9): 839–843. [DOI]
- [25] Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li YQ, Fine EJ, Wu XB, Shalem O, Cradick TJ, Marraffini LA, Bao G, Zhang F. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 2013, 31(9): 827–832. [DOI]
- [26] Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. *Nat Methods*, 2014, 11(2): 122–123. [DOI]
- [27] Lei Y, Lu L, Liu HY, Li S, Xing F, Chen LL. CRISPR-P: a web tool for synthetic single-guide RNA design of CRISPR-system in plants. *Mol Plant*, 2014, 7(9): 1494–1496. [DOI]
- [28] Montague TG, Cruz JM, Gagnon JA, Church GM, Valen E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res*, 2014, 42(Web Server issue): W401–W407. [DOI]
- [29] Naito Y, Hino K, Bono H, Ui-Tei K. CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites. *Bioinformatics*, 2015, 31(7): 1120–1123. [DOI]
- [30] Xiao A, Cheng ZC, Kong L, Zhu ZY, Lin S, Gao G, Zhang B. CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics*, 2014, 30(8): 1180–1182. [DOI]
- [31] Upadhyay SK, Sharma S. SSFinder: high throughput CRISPR-Cas target sites prediction tool. *BioMed Res Int*, 2014, 2014: 742482. [DOI]
- [32] Xie SS, Shen B, Zhang CB, Huang XX, Zhang YL. sgRNAs9: a software package for designing CRISPR sgRNA and evaluating potential off-target cleavage sites. *PLoS One*, 2014, 9(6): e100448. [DOI]
- [33] Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A,

- Enright AJ. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res*, 2006, 34(Database issue): D140–D144. [DOI]
- [34] Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, Durbin R, Eyras E, Gilbert J, Hammond M, Huminiecki L, Kasprzyk A, Lehvaslaiho H, Lijnzaad P, Melsopp C, Mongin E, Pettett R, Pocock M, Potter S, Rust A, Schmidt E, Searle S, Slater G, Smith J, Spooner W, Stabenau A, Stalker J, Stupka E, Ureta-Vidal A, Vastrik I, Clamp M. The Ensembl genome database project. *Nucleic Acids Res*, 2002, 30(1): 38–41. [DOI]
- [35] Shen B, Zhang WS, Zhang J, Zhou JK, Wang JY, Chen L, Wang L, Hodgkins A, Iyer V, Huang XX, Skarnes WC. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods*, 2014, 11(4): 399–402. [DOI]
- [36] Tsai SQ, Wyvekens N, Khayter C, Foden JA, Thapar V, Reyon D, Goodwin MJ, Aryee MJ, Joung JK. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol*, 2014, 32(6): 569–576. [DOI]
- [37] Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res*, 2003, 31(13): 3406–3415. [DOI]
- [38] Ambros V. microRNAs: tiny regulators with great potential. *Cell*, 2001, 107(7): 823–826. [DOI]
- [39] Xie SS, Li XY, Liu T, Cao JH, Zhong Q, Zhao SH. Discovery of porcine microRNAs in multiple tissues by a Solexa deep sequencing approach. *PLoS One*, 2011, 6(1): e16235. [DOI]
- [40] Zhao YC, Dai Z, Liang Y, Yin M, Ma KY, He M, Ouyang HS, Teng CB. Sequence-specific inhibition of microRNA via CRISPR/CRISPRi system. *Sci Rep*, 2014, 4: 3943. [DOI]
- [41] Jiang Q, Meng X, Meng LW, Chang NN, Xiong JW, Cao HQ, Liang ZC. Small indels induced by CRISPR/Cas9 in the 5' region of microRNA lead to its depletion and Drosha processing retardance. *RNA Biol*, 2014, 11(10): 1243–1249. [DOI]
- [42] 刘宣, 季青, 李文, 周利红, 王璇, 江海丽, 陈静, 畅立圣, 李琦. 应用 CRISPR/Cas9 基因编辑技术建立 miRNA-301a 敲除小鼠. *第二军医大学学报*, 2015, 36(3): 256–260. [DOI]
- [43] Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol*, 2014, 32(12): 1262–1267. [DOI]

(责任编辑: 吴强)